Higher-Order Networks

So Much Types of Networks

Networks help in **representing** and **analysing complex systems**

Chosen **network model depends on the data**

Temporal Networks, Bipartite Networks, Weighted…

Sequential Data and Networks

What about Pathway/Sequential networks ?

Standard network model depicts **pairwise interactions**

… sometimes simple models are not meaningful

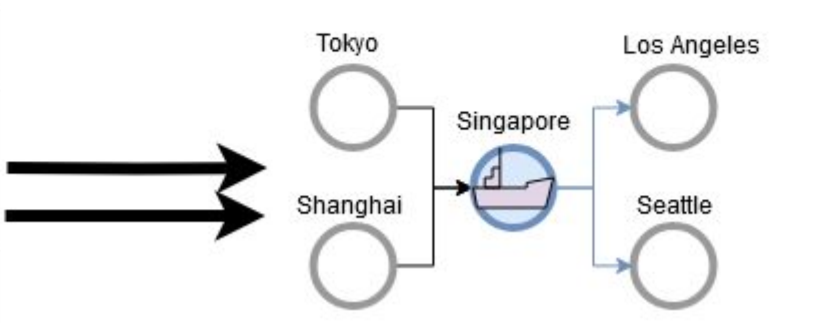
Sequential Data and Networks

What about Pathway/Sequential networks ?

Standard network model depicts pairwise interactions

<u>… sometimes simple models are not meaningful</u>

Vessel	Departure	Sailing data	Arrival	Arrival Date
v1	Shanghai	2013/01/01	Singapore	2013/01/15
v1	Singapore	2013/01/16	LA	2013/02/05
v2	Singapore	2013/02/01	LA	2013/03/08

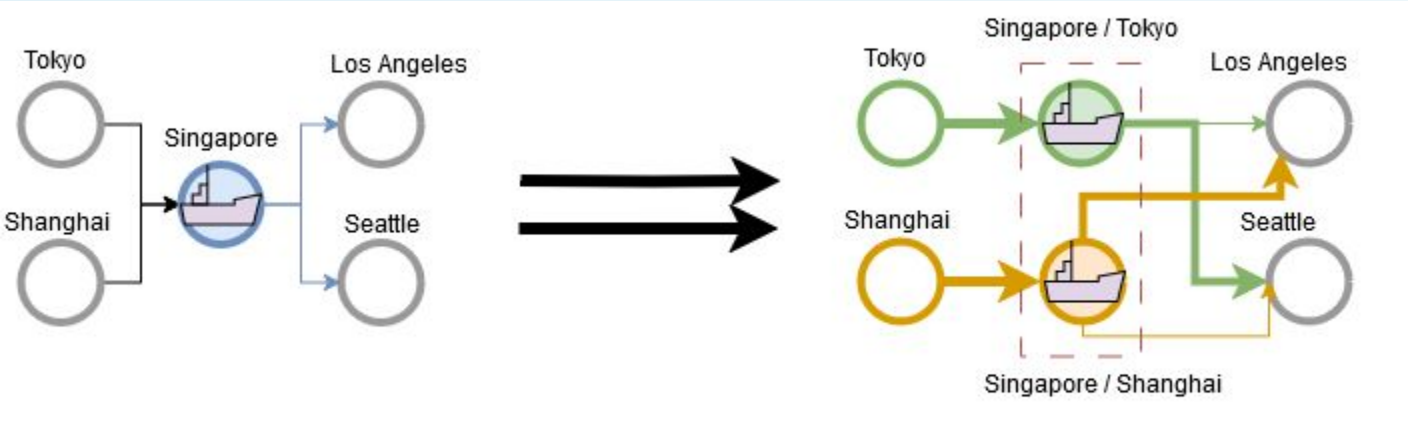


Xu et al. 2016. Representing higher-order dependencies in networks

Sequential Data and Networks

In case of trajectories for instance, **ordering**, long range and **memory of previous states** is **important**

Vessels can have **particular destinations** based on **initial ports**



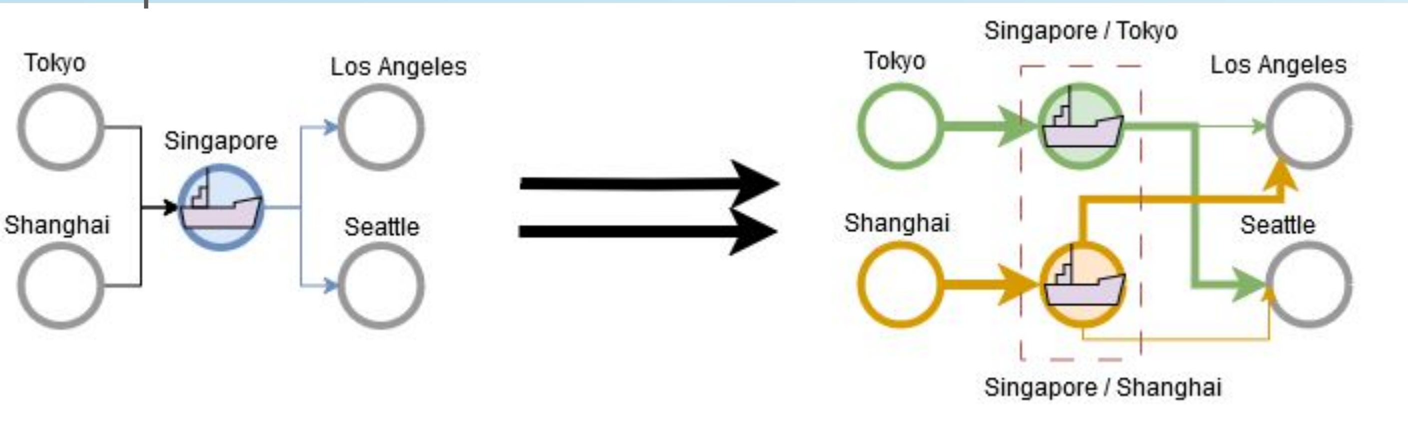
Xu et al. 2016. Representing higher-order dependencies in networks

Sequential Data and Networks

We need to go beyond pairwise interactions

First-Order Representation

Higher-Order Representation



Xu et al. 2016. Representing higher-order dependencies in networks

What are Higher-Order Networks?

Higher-Order Network (<u>HON</u>) = **Network representation**

Nodes represent **sequence of states** (<u>memory nodes</u>)

Edges represent **transitions** to possible **next states** (transition probabilities)

Random Walk through HON = **Non-Markovian process!**

Many real-world applications: Biological pathways, trajectories, urban traffic, social interactions

From Sequences to HON

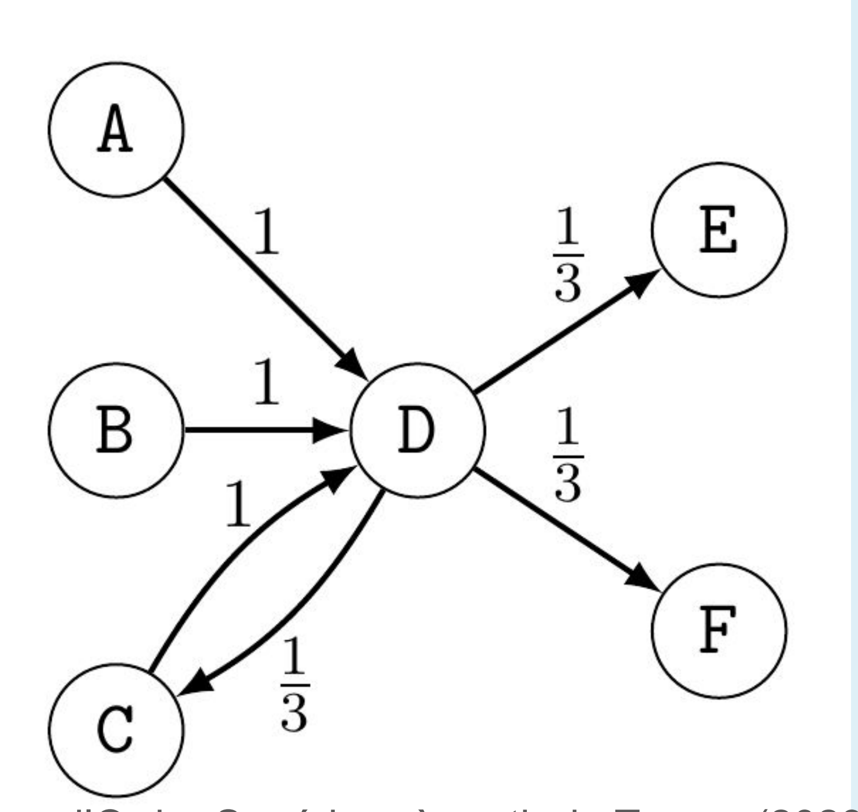
Séquences	#
A D E	2
D E	1
B D E	3
D F	2
B D F	4
B D C	3
C D C	3

Queiros and Queyroi, Construction de Réseaux d'Ordre Supérieur à partir de Traces (2023)

Direct Dependencies are Poor

First-Order Network

Séquences	#
A D E	2
D E	1
B D E	3
D F	2
B D F	4
B D C	3
C D C	3

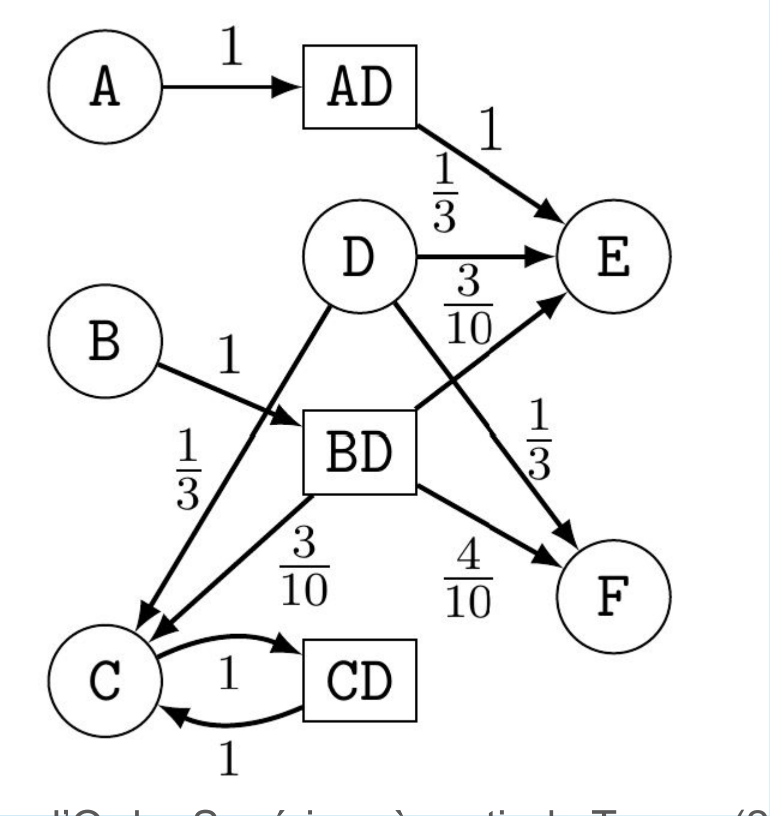


Queiros and Queyroi, Construction de Réseaux d’Ordre Supérieur à partir de Traces (2023)

Looking In The Past

Séquences	#
A D E	2
D E	1
B D E	3
D F	2
B D F	4
B D C	3
C D C	3

Second-Order Network

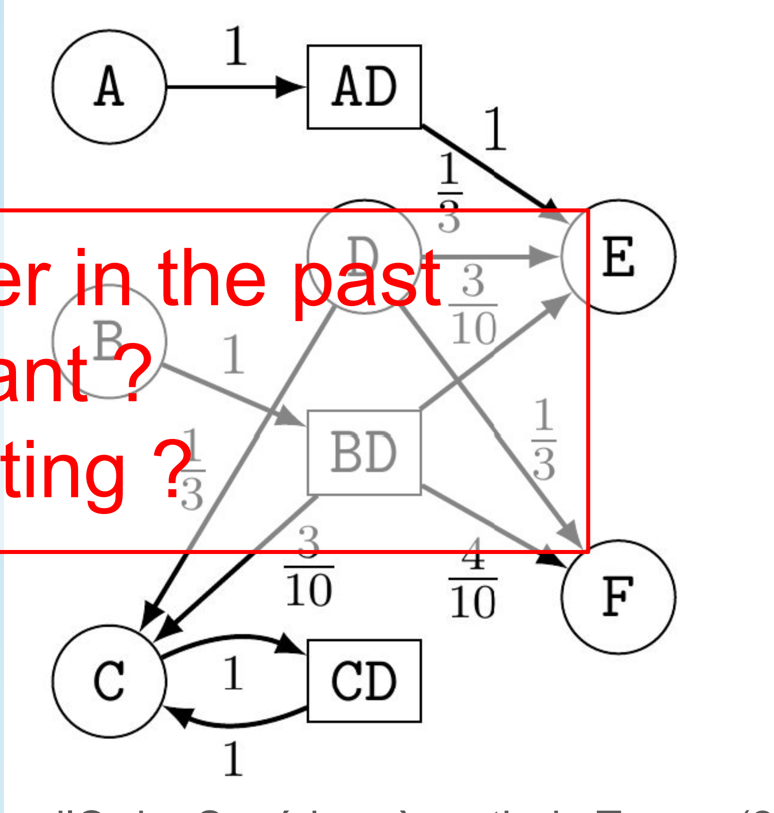


Queiros and Queyroi, Construction de Réseaux d’Ordre Supérieur à partir de Traces (2023)

Keep Looking back?

Séquences	#
A D E	2
D E	1
B D E	3
D F	2
B D F	4
B D C	3
C D C	3

Second-Order Network



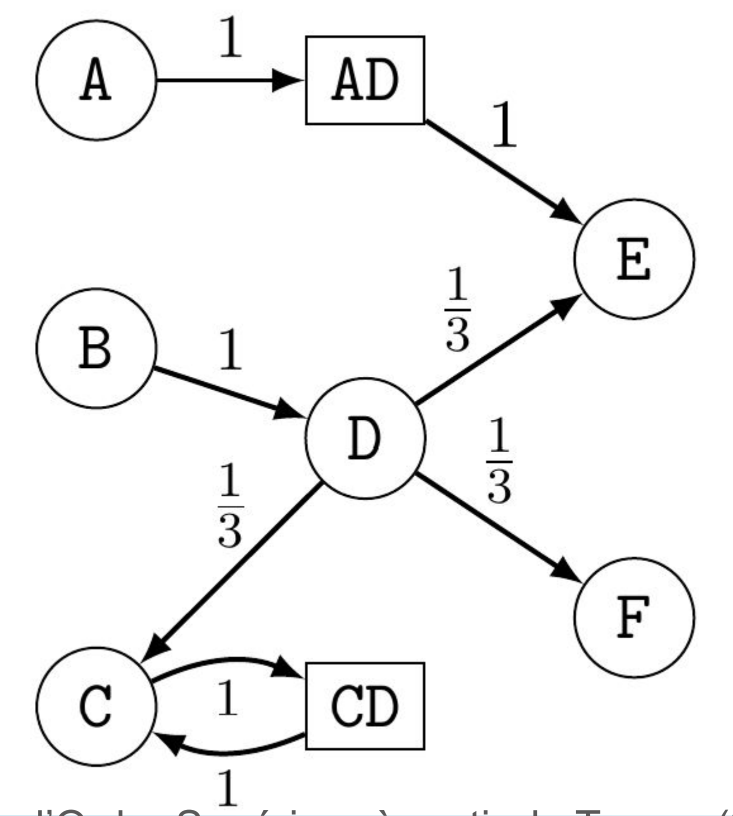
Is going further in the past relevant ?
Overfitting ?

Queiros and Queyroi, Construction de Réseaux d'Ordre Supérieur à partir de Traces (2023)

Not All Past is Relevant

Séquences	#
A D E	2
D E	1
B D E	3
D F	2
B D F	4
B D C	3
C D C	3

Variable-Order Network



Queiros and Queyroi, Construction de Réseaux d'Ordre Supérieur à partir de Traces (2023)

Pros and Cons of HONs

+ Capture **long range interactions** between **sequence elements**
+ Walk through HONs give **better sequence predictions**

- **Larger models**
- Require **sequence mining** + counting (<u>computer cost</u>)
- **What is a node**? **Maximum order**?

Definitions

Let A be the <u>set of all possible states</u>

S is the <u>multi-set of sequences</u> s_i

with $s = \sigma_1\sigma_2\ldots\sigma_m$ an <u>ordered sequence of states</u>

<u>Order of *s*</u> is noted $|s|$ = **number of states**

c(*s*) is the <u>number of occurrence of *s*</u> in ***S***

Definitions

s' **suffix of** s iif $|s'|$ **last elements** of $s = s'$

s' **prefix of** s iif $|s'|$ **first elements** of $s = s'$

Depending on the sequential model s can be (statistically) represented by a suffix s' called context

{AD, ADE, DF} <u>included in</u> ***S***

{AA, DEDF} <u>are not</u>

E is a **1st-order** seq
c(E) = 6

BD is a **2nd-order** seq
c(BD) = 10

Séquences	#
A D E	2
D E	1
B D E	3
D F	2
B D F	4
B D C	3
C D C	3

p(BD -> E) = p(E|BD) = ?

Séquences	#
A D E	2
D E	1
B D E	3
D F	2
B D F	4
B D C	3
C D C	3

p(BD -> E) = p(E|BD) = ?

$$P(E|BD) = \frac{c(BDE)}{\sum_{\sigma \in A} c(BD\sigma)}$$

= 3/10

Séquences	#
A D E	2
D E	1
$\boxed{\text{B D E}}$	$\boxed{3}$
D F	2
B D F	4
B D C	3
C D C	3

HONs are Stochastic Graphs

G = (V,E)

V = Set of **memory nodes** being **sequential contexts**

a **node *v*** is a **representation of a state** from ***A***

Ex: <u>AD </u>is a <u>representation of the state D</u>

"Arrived at D with A as previous state"

E = Set of **directed** and **weighted edges**

$P(\sigma|s) > 0$ associated with $s \to s^*\sigma$

Fixed-Order Network Models

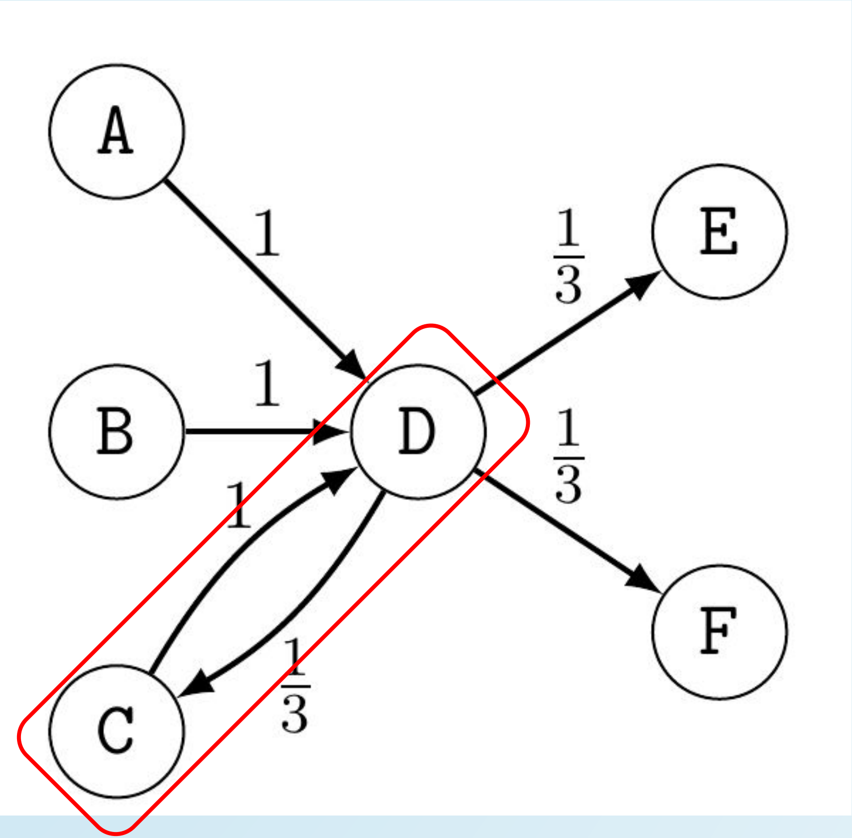
Fixed-Order Network (FON)

First-order network ($\underline{FON_1}$)

Contexts of **size 1**

$\underline{P(C|CD)} = \underline{P(C|D)} = 1/3$

P(.|D) = [0,0,1/3,0,1/3, 1/3]

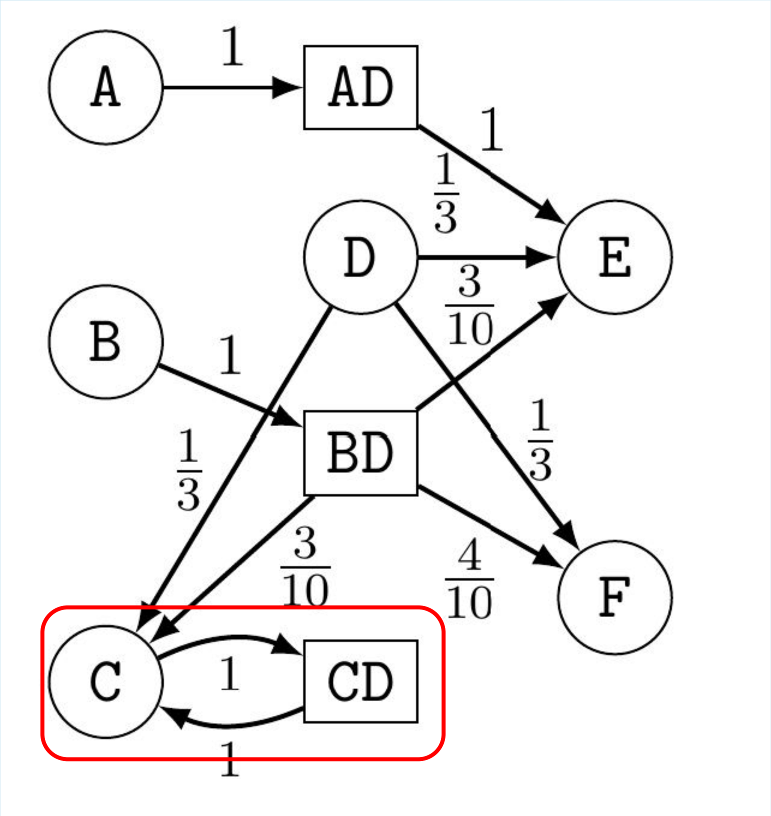


Fixed-Order Network (FON)

Second-order network $\underline{FON_2}$

Contexts of **size 1** and **2**

$\underline{P(C|CD) = 1}$



Fixed-Order Network (FON)

Generalisation to **FON**$_k$ possible

… but **becomes very large** $\mathcal{O}(N^k)$

… and **which *k*?**

FON_k construction process

1 **Contexts** are all **sequences** of **size <= *k***

2 **Adding nodes** and **edges** from considered **contexts c** and **distributions P(.|c)**

Ex: "ADE" in FON_2 gives A -> AD -> E

"DE" gives D->**E** (**E is a stop here**)

Variable-Order Network Model

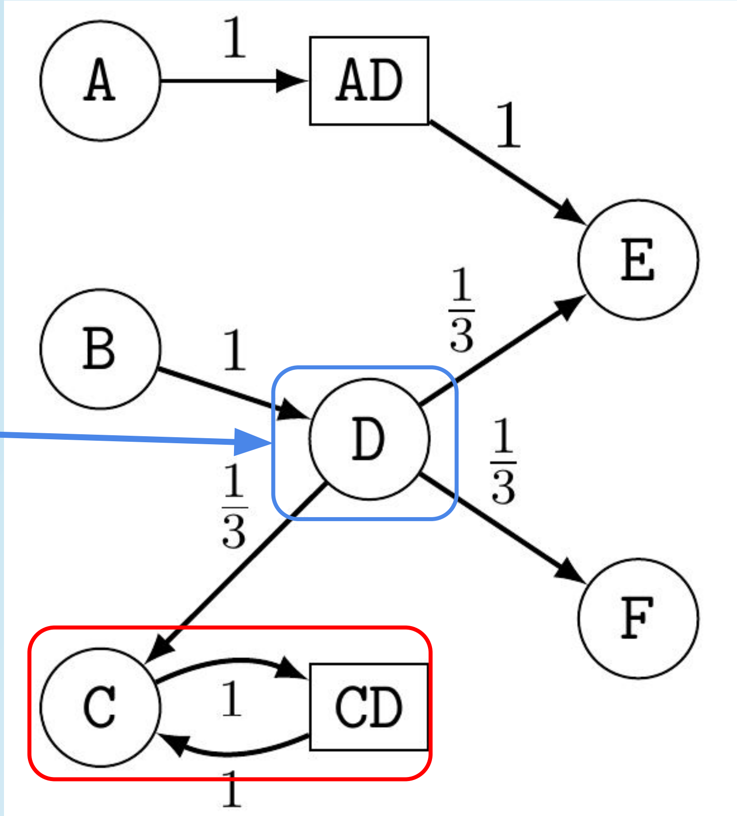
Variable-Order Network (VON)

Contexts of **variable size**

P(C|CD) = 1

Keep only **informative contexts**

Ex: P(.|BD) similar to P(.|D)



Variable-Order Network (VON)

Context of **different sizes** are considered

Statistically relevant contexts are kept

Use of **quality function**!

The **Kullback-Leibler divergence** D_{KL}

Variable-Order Network (VON)

The Kullback-Leibler divergence D_{KL} is a **measure used** in **information theory**

For a context s, **let Q** be the **distribution** $P(.|s)$ and **P** the **distribution** associated to s' **an extension of s** (s=suffix of s')

D_{KL} gives the additional information (bits) encoded by P regarding Q

$$D_{KL}(P||Q) = \sum_{\sigma \in A} P(\sigma) log_2(\frac{P(\sigma)}{Q(\sigma)})$$

Variable-Order Network (VON)

Context **relevant if D_{KL} > threshold**

Preference in keeping a **short** and **highly observed** sequence

$$D_{KL}(P||Q) > \frac{|s'|}{log_2(1 + c(s'))}$$

VON construction process

1 **Starts** with **first-order context** (elements from ***A***)

2 Context **extension** and $\mathbf{D_{KL}}$ **test**

3 **Adding nodes** and **edges from** considered contexts c and **distributions P(.|c)**

Reminders on HONs

FON	VON	Feature
Fixed *k*	Variable	Memory Length
Higher	Lower	Model Complexity
Limited	Better	Scalability

Many other models exist: Tensor based, multi-order network, von-sample..

Higher-Order Network Analysis

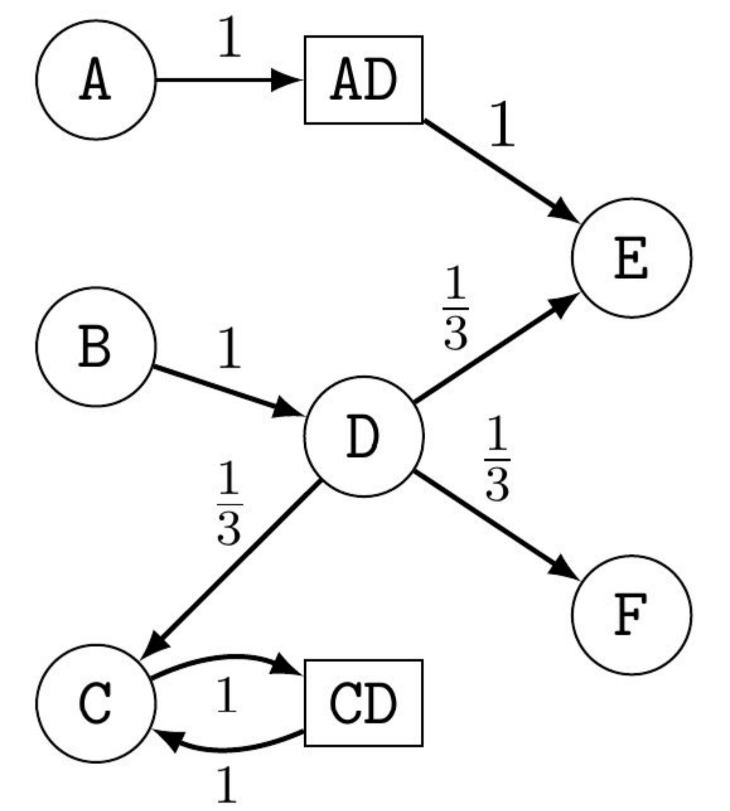
Analysing HONs

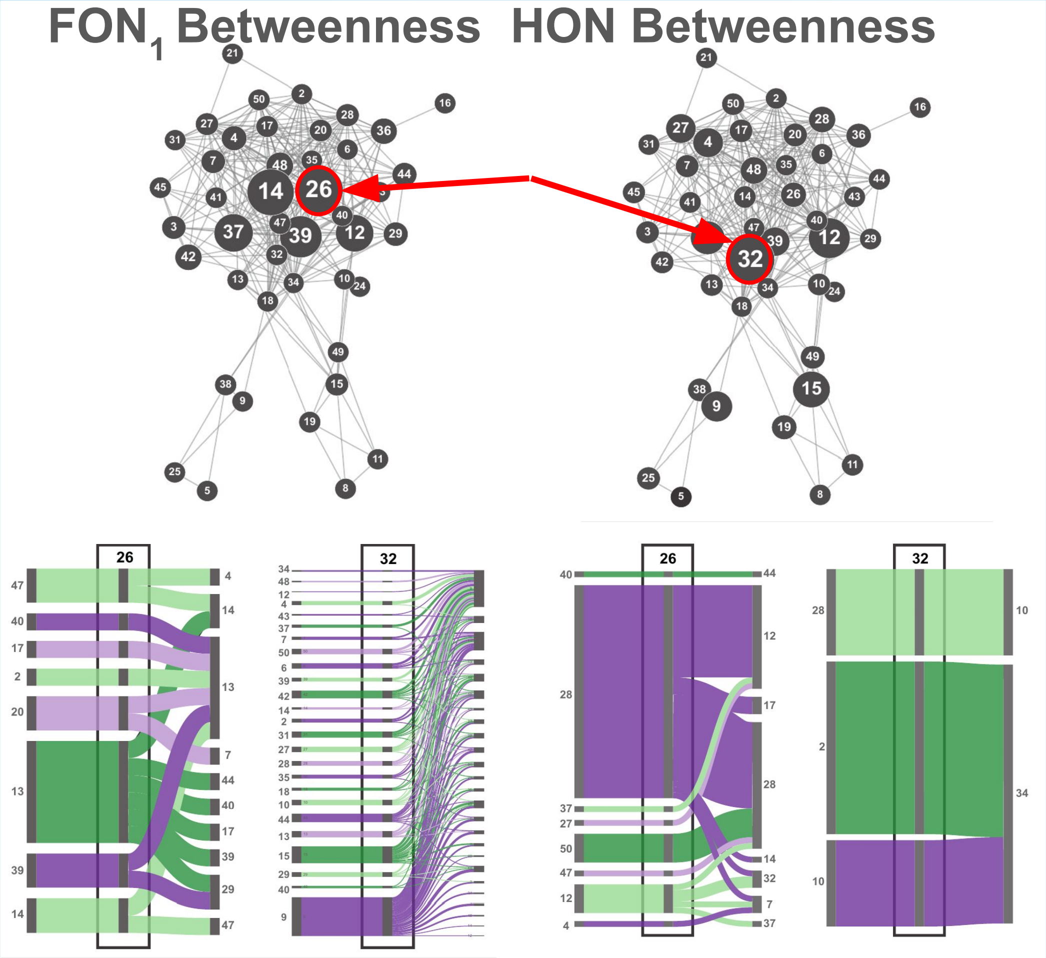
Are standard network analysis tools efficient for HONs?

What is a **clique in HONs**?

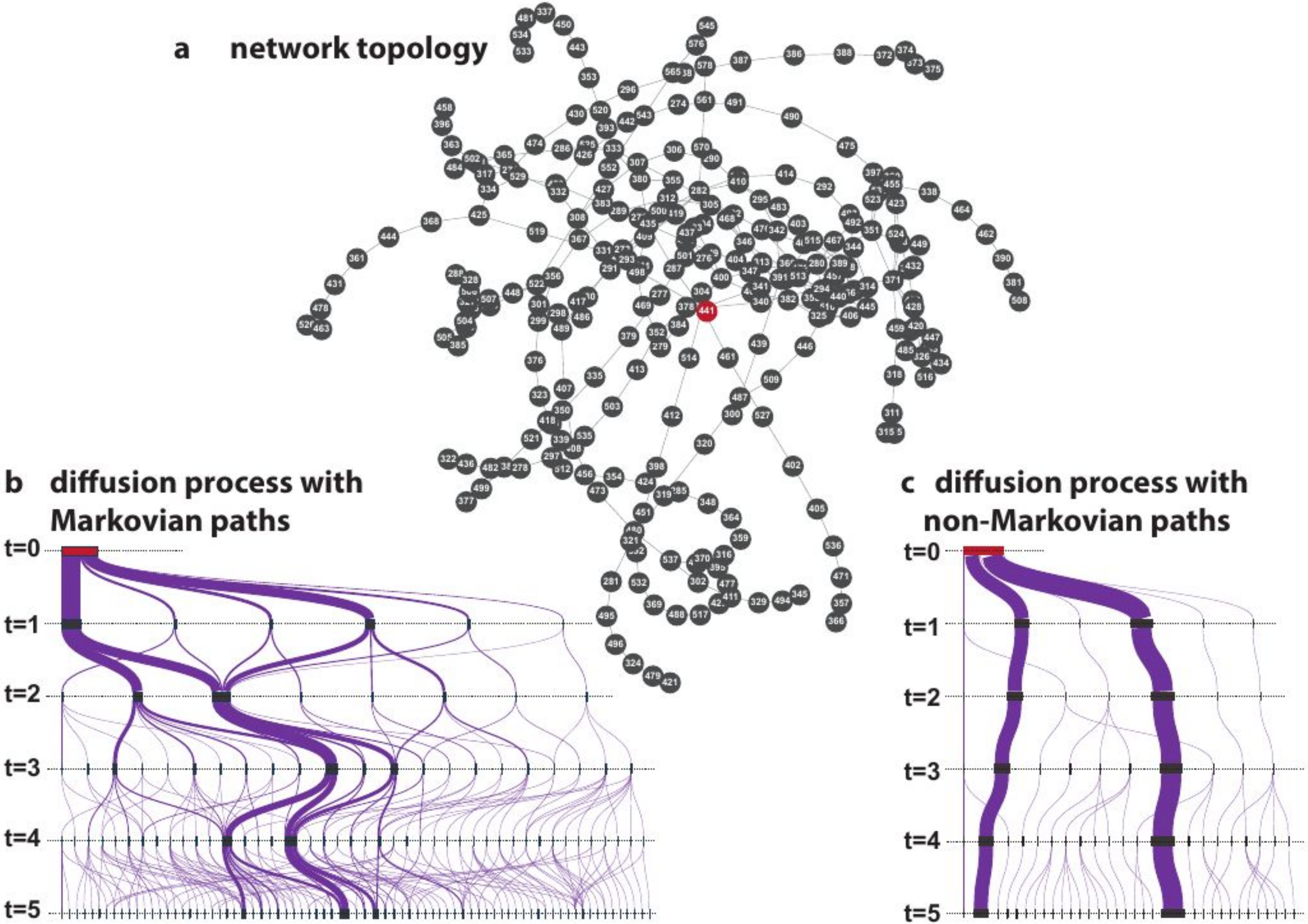
Degree?

Topology? **Centrality**?





Lambiotte et al. 2018. Understanding Complex Systems: From Networks to Optimal Higher-Order Models



Lambiotte et al. 2018. Understanding Complex Systems: From Networks to Optimal Higher-Order Models

PageRank on HONs

PageRank on HONs

HONs = Good network model

=> **Analysis Tools** applied on HONs would **give better results!**

(... true ?)

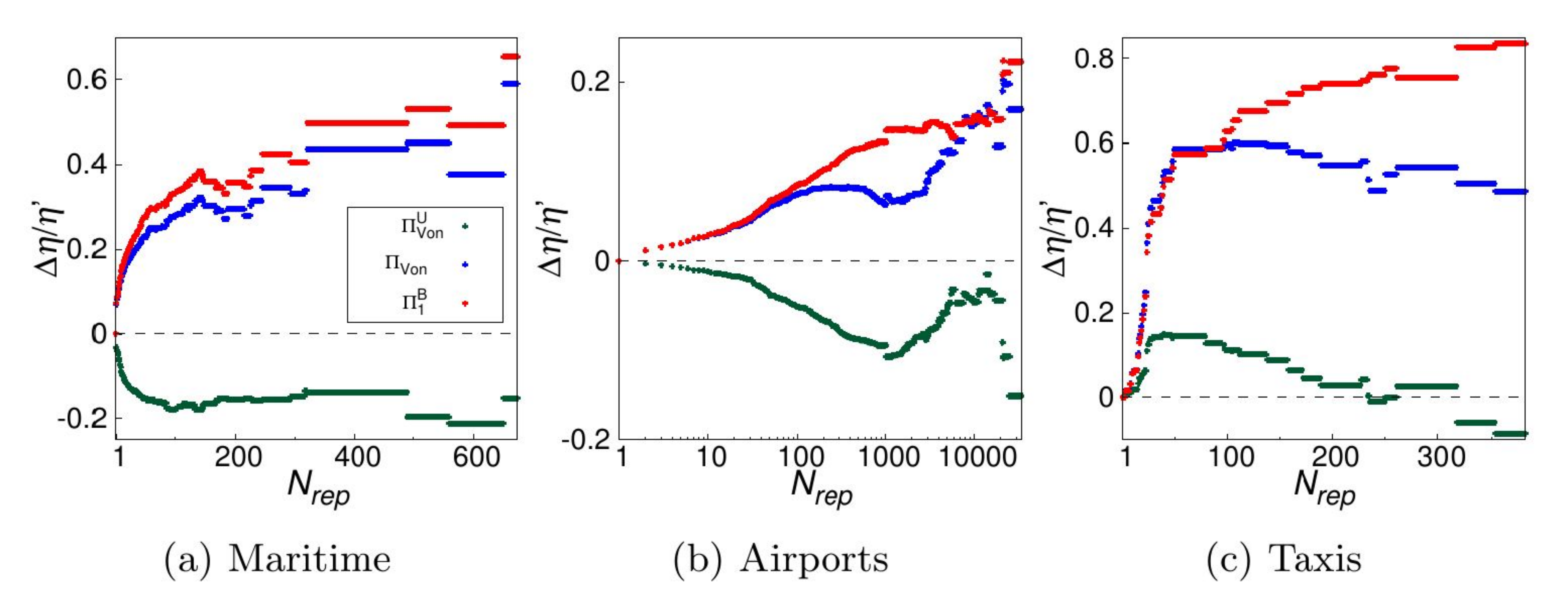
PageRank on HONs

Score of **state** σ is **calculated from its representations' scores**

In case of **PageRank**, **highly represented states** can **lead to biases**

Because of **standard PageRank** algorithm the **more a state is represented** the more its **representations will be visited** (due to **teleportation)**

Need to **adapt** the tool **to HONs**



Coquidé et al. 2021. PageRank computation for Higher-Order Networks

Additional Notes

The **term Higher-Order** Network is **highly used in the literature**

Most of the research on **pathway networks** use it, but **other usage too**

Among them we have:

- **Hypergraphs** (<u>edge between more than 2 nodes</u>)
- **K-cliques** analysis
- **Simplicial complex**